DOCUMENT RESUME

ED 250 747 CS 504 759

AUTHOR Warfel, Katherine Ann
TITLE Use of the Rasch Model in Communication Education: An Explanation and Example Application.
PUB DATE Nov 84
NOTE 27p.; Paper presented at the Annual Meeting of the Speech Communication Association (70th, Chicago, IL, November 1-4, 1984).
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Research; *Item Analysis; *Latent Trait Theory; Measurement Techniques; *Models; *Speech Communication; Speech Instruction; *Test Construction; Testing; Test Items; *Test Theory; Test Validity
IDENTIFIERS *Rasch Model

ABSTRACT

The goal of test design is to devise an instrument that will provide a stable and accurate assessment of student ability in some area. One means of reaching this goal is through the use of latent trait models, which determine the relationship between the unobservable trait or ability and the observable test performance. Three common latent trait models include the one parameter (or Rasch) logistic model, the two parameter logistic model, and the three parameter logistic model. Of these, the Rasch model is preferred, due to its ease of application and the fact that test scores serve as a sufficient statistic for ability estimation. In addition, computer programs are more widely available for the Rasch model than for the other two models. Latent trait models also have a number of characteristics that make them superior to classical test models. Among them are the properties of objectivity and information, and the ability to identify biased test items. To test its application in the context of communication education, the model was used with two parallel forms of a 75-item multiple choice final examination. Given the results of the initial factor analyses, the data fit the assumption of test unidimensionality, as required by the Rasch model analysis. The model was able to recover known item parameters ("known" in the sense that they should be equal), and the shapes of the two test information functions were bell-shaped rather than flat. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED250747

Use of the Rasch Model in Communication Education: An Explanation and Example Application

Paper submitted to the Instructional Development Division

Speech Communication Association Convention

November, 1984

Chicago, Illinois

Katherine Ann Warfel, M.A.

University of Texas at Austin, TX

Department of Speech Communication

Austin, Texas 78712

512/471-4130

512/471-5251

CS504759

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Katherine A. Warfel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

A recent development in testing theory is latent trait modeling. These models are superior to classical test models, yet have not realized significant use in the field of communication education. The purpose of this paper is to provide a brief, non-technical overview of latent trait models, as well as to provide an argument for the superiority of latent trait models (and the Rasch logistic model, in particular) over classical test models. Finally, an example application of the Rasch model in the context of communication education is demonstrated.

Use of the Rasch Model in Communication Education:

An Explanation and Example Application

An emerging trend in education is the movement of students from smaller, privately funded colleges to larger, publicly funded state universities. With current enrollments at these universities in the tens of thousands, it is not unusual to find 500 or more students registered for introductory level lecture courses in a variety of disciplines. This growth trend in class size has impacted upon the field of communication education as well. Although many universities maintain a much smaller limit on the so-called "performance" classes (e.g., public speaking); numerous schools have instituted the lecture format for a number of introductory communication theory courses (e.g., interpersonal or organizational communication) in order to accomodate large numbers of students.

Given class sizes of this magnitude, many educators have been forced to abandon traditional "blue book" essay exams in favor of the more efficiently graded multiple choice objective text. With the advent of these tests have come a number of concomitant problems. To enforce test security, educators often develop several forms of any one exam and/or change test forms completely from one semester to the next. This practice raises several testing issues. Are different tests and/or test forms of equal levels of difficulty? Are different tests and/or test forms tapping the same body of knowledge or skills? Students often complain about the irrelevance and unfairness of "picky and ambiguous" test questions. How can a "good" test

item be distinguished from a "bad" test item? How can one identify test items that discriminate against certain groups? Often, a number of parallel sections of a single course will receive the same departmental course exam. How can a test be constructed that will provide fair and accurate comparisons of student performance across parallel course sections? These are some of the thorny issues facing multiple-choice test designers. The purpose of this paper is to discuss a valuable tool for coping with these difficult problems.

The goal of testing is to devise some instrument that will provide a stable and accurate assessment of student ability in some area. One means of reaching this goal is through the use of latent trait models. Latent trait models are useful in test design and are helpful in solving testing problems that had seemed difficult to resolve. They have, however, seen relatively little use in the field of communication education. This is probably due to several reasons. First, most of the early work with latent trait models has been done in the field of test theory. The mathematics presented in the early papers on the subject may have seemed formidable to those outside the field. Second, most of the work in this area has been directed to theory, rather than practice. It is difficult for those without a background in test theory to make practical use of the models on their own. Finally, until recently, computer programs that quickly and easily estimate the parameters used in the models were not widely available. This is no longer true, especially in the case of the Rasch, or one parameter, logistic model (Hambleton and Cook, 1977).

The purpose of this paper is to provide a brief, non-technical overview of latent trait models and their application to communication education. Latent trait models will be compared to classical test models, with an argument made for the superiority of latent trait models for testing purposes. Finally, an example of the use of the Rasch model (one of the latent trait models) in the context of communication education will be provided.

The Historical Development of Testing

The first recorded uses of testing were in China over 3000 years ago. Through the years, tests have been used for a number of purposes: to assess progress and/or achievement in education, for selection of job candidates, and so on. Current interest in testing may be traced to three sources: civil service exams, individual difference testing, and educational testing (Allen and Yen, 1979).

Civil service examinations were present in ancient China and emerged sometime later in Great Britain and the United States. They became popular as a means of removing job candidate selection from political bias. Individual difference testing began in earnest in the late 1800's with the study of human sensory-motor skills by Sir Francis Galton. Later, Binet and Stern concentrated on measures of intelligence. Researchers such as Pearson and Spearman helped to develop the sophisticated techniques used today. Educational testing was in the form of oral examination until the use of paper was introduced after the twelfth century. Much later, in the sixteenth century, Jesuits began the use of testing to evaluate their students (Allen and Yen, 1979).

Of course, all are familiar with the widespread use of testing in the modern educational systems. It is to the development of measures used in the assessment of such tests that attention is now turned.

Early tests were usually administered to one person at a time. The examiner would then make some subjective decision about the testee's performance on the test. Later, especially during and after World War I, the large scale use of group testing became more common. With the widespread use of such tests for placement and evaluation purposes, it became necessary to devise adequate measures to assess the validity of these tests. One major concern has been the derivation of scales used for placement and evaluation of examinees. This process has been criticized in that scales derived from the performance of some sample of testees might not be generalizable across larger examinee populations. Indeed, until recently, procedures did not exist to develop "sample free" scales. In his work with intelligence tests for the Danish Department of Defense in 1945, Georg Rasch pointed out the additional problem of accurately defining the ability of an individual independently of the set of test items the individual actually solved (Wright and Stone, 1979). Although it is not possible to circumvent these issues with traditional testing methods; latent trait models (and the Rasch model in particular) do provide the means to solve these testing problems.

Latent Trait Models

In this section, characteristics of latent trait models will be discussed. Types of latent trait models will be described

and an argument made for the use of the Rasch model over other latent trait models. Finally, a discussion of the superiority of latent trait models over classical test models will be presented.

Simply stated, latent trait models are based on the notion that there is some unobservable trait or ability possessed by all testees to some degree (hence, the name "latent"). Further, it is assumed that test performance can be explained by the presence or absence of this trait or ability. The latent trait model determines the relationship between the unobservable trait or ability (e.g., knowledge of interpersonal communication theory) and the observable test performance (e.g., testees' grade on an exam). This relationship is illustrated by a mathematical function. Therefore, latent trait models may be thought of as mathematical models (Hambleton and Cook, 1977).

There are several important characteristics of latent trait models. The first is an assumption regarding the number of latent traits underlying testees' performance on some examination. Recall from above that each examinee's performance (or "latent trait score") is based on the presence or absence of a latent trait or ability. It is typically assumed that a given test taps only one ability or trait at a time. That is, items are assumed to be homogeneous to the extent that they all measure one ability. For example, a vocabulary test taps a certain verbal ability; a test on multiplication taps a certain mathematical ability, and so on. Furthermore, a person's performance on one test item is assumed to be statistically independent of his/her performance on other items in the test. In effect,

this is a correlate of the above in that one is again stating (although in a slightly different way) that performance on test items is governed by one single ability (Hambleton and Cook, 1977).

This notion of single ability testing, or unidimensionality, may be investigated through the use of factor analytic methods. One simply factor analyzes the dichotomous (right/wrong) scores of the examinee population. If the first factor is large compared to the second; and the second factor is not relatively large compared to the third and lower factors; then one can reasonably assert that one's test is unidimensional (Lord, 1980). If several distinct factors emerge, then one may cluster the homogeneous items that load on each orthogonal factor and treat these sub-tests as unidimensional for the purposes of the latent trait model analysis (Hambleton and Cook, 1977).

The second feature of latent trait models is that of the item characteristic curve. This curve is a mathematical function that relates the probability of success on the test item to the ability level of the testee. These curves will vary, depending on the latent trait model one chooses to utilize. It is assumed that the probability of a testee answering an item correctly is independent of that testee's ability relative to others in the testee population. In this sense, the shape of the curve is <u>not</u> dependent on the distribution of the ability of the latent trait in the population of examinees (Hambleton and Cook, 1977). The importance of this feature will become clear in the discussion of "sample free" item calibration presented a bit later in this paper.

Types of Latent Trait Models

There are three common latent trait models: the one parameter (or Rasch) logistic model, the two parameter logistic model, and the three parameter logistic model. These models differ in the number of parameters taken into account by the latent trait model analysis. The two and three parameter models will be discussed first, followed by a more extensive treatment of the one parameter, or Rasch, logistic model. For a more complete treatment of the mathematical proofs and procedures used in these models, see Allen and Yen (1979).

Two and three parameter logistic models. The two parameter logistic model produces an item characteristic curve that takes into account both item difficulty and item discrimination. Item difficulty is measured at the point on an ability scale at which the slope of the item characteristic curve reaches a maximum value. Item discrimination refers to an item's power to distinguish between different levels of testee ability. Item discrimination is measured proportional to the steepness of the slope of the curve (Allen and Yen, 1979; Hambleton and Cook, 1977).

The three parameter logistic model produces an item characteristic curve that takes into account both of the parameters cited above, plus a parameter that estimates guessing by testees. That is, the third parameter represents an adjustment of the lower asymptote of the curve to deal with the possibility and amount of persons with low ability correctly guessing the answer to an item (Allen and Yen, 1979; Hambleton and Cook, 1977).

Debate continues over the validity of these two models. Wright and Stone (1979) argue that use of the two and three

parameter models is problematic in that:

> . . . only item difficulty can actually be estimated consistently and sufficiently from the right/wrong item response data available for item analysis. These proofs make it clear that the dichotomous response data available for item analysis can only support the estimation of item difficulty and that attempts to estimate any other individual item parameters are necessarily doomed (p. ix).

For this reason, many argue for the superiority of the one parameter, or Rasch, model over the two or three parameter models. In addition, the Rasch model is easier to work with than the other logistic models, as only one parameter must be calculated. The Rasch model has other characteristics that make it more attractive than the other logistic models. These will become clear in the discussion of the Rasch model that follows.

The Rasch model. The Rasch model, or one parameter logistic model, takes into account only item difficulty. Item discrimination is assumed to be equal across items, and testee guessing is assumed to be non-existent. Hambleton and Cook (1977), among others, contend that these assumptions are highly restrictive, and call into question the robustness of the model. Wright (1967, 1977) argues that, with a well constructed test, guessing is minimized. Also, prior examination of items should lead to a final pool of items with equal discrimination. In this case, the impact of the restrictions is minimized. In addition, Wright offers evidence that the model is robust, even with data that does not "fit"

the assumptions of the model (Wright, 1967).

In addition to the ease of application of the Rasch model, there is another advantage to its use. With the Rasch model, a person's test score is a "sufficient statistic" to estimate ability. That is, given a person's score, one needs no other parameters in order to estimate that person's ability. This is a distinct advantage over the other two models, where one must also estimate measures of discrimination and level of guessing before one can produce a measure of testee ability. As was mentioned earlier, calculation of these parameters may be problematic (Hambleton and Cook, 1977; Wright, 1977).

In summary, it would seem that, given scores from a well constructed test, use of the Rasch model is advantageous due to its ease of application and the fact that test score serves as a sufficient statistic for ability estimation. In addition, computer programs are more widely available for the Rasch model than for the other two models.

Superiority of Latent Trait Models

There are three major reasons for the superiority of latent trait models over classical test models. The first is the property of "objectivity" possessed by latent trait models. The second is the property of "information" found in the item characteristic curve. The third is the ability to determine the presence of biased items in a test. Each of these properties and their importance will be discussed.

First, as was mentioned earlier, test scales based on the performance of one sample of examinees were criticized as being nongeneralizable across other samples of examinees. For example,

an items point biserial (PBIS on the standard item analysis printout) is a measure of correlation between persons who answered an item correctly and those who achieved the highest scores on the test overall. However, this measure is dependent on the sample of persons who took that particular test, and will vary across examinee samples and across tests. In addition, Rasch (1966) has pointed out the difficulty in estimating testee ability independently of the items s/he actually solved. Latent trait models resolve these issues through the property of "objectivity". According to Wright (1967), two conditions are necessary to achieve objectivity. First, instrument calibration must be object free; and second, object measurement must be instrument free. This is achieved in latent trait models through "sample free" item calibration and "item free" person measurement (Hambleton and Cook, 1977; Wright, 1967, 1977).

"Sample free" item calibration refers to the advantage of latent trait models that item parameters do not vary across samples, as is the case with classical test parameters. For example, "P-values" obtained from a standard item analysis are sample dependent in that they represent the proportion of those in the sample of examinees who answered the item correctly. Rasch item difficulties, on the other hand, correct for mean ability in the sample and for sample variance. For this reason, Rasch item difficulties are invariant across samples or across time (Wright, 1977; Wright and Stone, 1979).

"Item free" person measurement refers to the quality of latent trait models that, once item parameters are established, a person's ability may be accurately estimated from any subset

of these items (Hambleton and Cook, 1977; Wright, 1977). The advantages of this preperty are numerous. Given a large pool of pre-calibrated items, one could construct a large number of different, but parallel, test forms. This could solve the problem of test security, while at the same time providing comparable estimates of student ability across sub-samples of items. In addition, the Rasch ability estimate is superior to the "score" a person receives on a traditional test because it has been transformed to make the estimate a test-free measure (Wright and Stone, 1979). This makes it possible to make meaningful comparisons of persons' scores across different examinee samples, different test forms, and across time. This is not possible with "scores" obtained from classical test models.

The second property that makes latent trait models superior to classical test models is that property of "information". In classical test models, only one standard error of measurement is produced for the entire sample of testees, regardless of their test score. In latent trait model analysis, it is possible to calculate separate standard errors of measurement for each level of ability. One can then calculate an item's information function, based on the slope of the item characteristic curve and the standard error of measurement for some specific ability level. The resulting figure gives an estimate of the accuracy of ability estimation at that point. Due to the precision of this method, it has been argued that test information curves are superior to traditional estimates of test reliability (e.g., alpha) (Hambleton and Cook, 1977).

A third advantage of latent trait models over classical models is the ability to identify test items that are biased against sub-samples of a group of examinees. Some items are presented in terms that may be difficult for some persons to comprehend. This may lead to performance on an item that is unrelated to the person's actual ability in that subject area. For example, a reading comprehension problem may contain terms that are unfamiliar to certain ethnic groups. This could result in an item that is biased against those ethnic groups. Through examination of the item characteristic curves, partialled across the groups in question, latent trait models provide a way to identify biased items, as well as producing quantitative support for removing or adjusting them (Lord, 1977; Wright, 1977). Documentation of this kind is not available with classical testing methods. In a time when discrimination and affirmative action are such sensitive issues, this feature of latent trait models becomes very attractive. (For a detailed explanation of the identification of bias in test items, see Lord (1977).)

In summary, latent trait models have a number of characteristics that make them superior to classical test models. These are: a) the property of "objectivity", b) the property of "information", and c) the ability to identify biased test items. In the following section, an example application of the Rasch, or one parameter, logistic model will be presented.

Example Application of the Rasch Model

In preceding sections, latent trait models were described and an argument made for the superiority of latent trait models over classical test models. In addition, the Rasch model was suggested as the most defensible of the latent trait models. In this section, one example of the application of the Rasch model in the context of communication education will be provided.

Purpose:

The purpose of the investigation was twofold: a) to test the Rasch model's ability to recover item parameters, and b) to evaluate the information functions of two parallel forms of a multiple choice test. The tests used in the investigation were two parallel forms of a 75 item multiple choice final exam. The exam covered the contents of a 16 week course in "Business and Professional Speaking", offered through a department of Speech Communication. Since the two forms of the test each consisted of the same 75 items (although presented in different orders), the item parameters (e.g., Rasch item difficulty) should be invariant across the two test forms. Also, as the two test forms are parallel, the test information functions should be nearly identical. In this sense, the first part of the investigation is a test of the Rasch model's ability to recover known parameters.

In the second part of the investigation, the shape of the information functions of the two tests will be evaluated. Since this is a test of course content, the tests should deliver reasonably equal amounts of information across all ability levles. That is, both test should generate reasonably flat information functions. In summary, the purpose of this investigation is twofold: a) test

the Rasch model's ability to recover known item parameters ("known" in the sense that they should be equal), and b) evaluate the two tests' information functions.

Sample:

The sample consists of 385 undergraduate students who took "Business and Professional Speaking" at the University of Texas in the Spring of 1983. The students (male and female) represent the typical age range for persons in this group--late teens to early twenties. 192 test forms were recovered from test form "A", while 193 test forms were recovered from test form "B". The tests were 75 items in length. Each question was in multiple choice format, with five choices each.

Methods:

Thr final exams were given to students in all sections of the course during the regular exam week. Care was taken to be sure that students sitting in close proximity to one another were given different forms of the test. Students used the standard computerized answer sheets to record their responses to the questions.

The first step in the process was to recover the individuals' computer test forms. Once this was accomplished, the two test forms were analyzed separately and standard item analyses generated. The response data from the two test forms was then converted to dichotomous (right/wrong) data through the use of a computer algorithm designed for this task. Once in this form, the data was factor analyzed to discern whether or not the test could be regarded as unidimensional--a key assumption of the latent trait model analysis. In the next step of the analysis, Rasch item difficulties were generated, using a computer algorithm constructed for this purpose.

The two sets of item difficulties were then correlated and a regression analysis performed on the two sets of figures. Finally, test information functions were generated using a computer program created for this function.

Results:

Factor analysis. The first step in the analysis was to determine if the tests were unidimensional in nature. That is, did the tests tap one range of ability, that of knowledge of the course content? Recall that test unidimensionality is an assumption of latent trati model analysis. The data from the two test forms was factor analyzed using principal factors analysis, varimax rotation, and a minimum eigenvalue of 1.5. According to Lord (1980), a test may be considered unidimensional if the first factor is relatively large compared to the second, and the second factor is large compared to the third and subsequent factors. In both the forms, the first factors accounted for over 24 percent of the variance in the data, while the second factors accounted for 7.8 percent and 8.6 percent, respectively. The third factors in both tests were similar in magnitude to the second factors, although all subsequent factors were quite small.

There is reason to call the validity of these last findings into question, however. Given that there were 75 items being factored on both tests, in order to achieve a stable solution there should have been at least 750 in each of the two samples (i.e., ten responses for each of the 75 items). As there were only approximately 200 of each of the two tests, one cannot expect a stable factor structure to emerge from the analysis. For this reason, it is difficult to rely on the accuracy of the analysis. Given the large magnitude of

the difference between the first set of factors and the subsequent factors, however, it seems safe to operate under the assumption that the tests are unidimensional.

Rasch item difficulties. The Rasch item difficulties from both test forms are presented in Table 1. Items on test form "A" run from a maximum "easiness" of +2.686 to a maximum "difficulty" of -3.079. On test form "B", the range is from +3.238 to -3.112. The distribution of the item difficulties is such that there is a reasonably even distribution of "difficult" and "easy" items on each test form. This is typical and desirable for tests of this kind.

The Pearson correlation coefficient was calculated to be +.907 (p = .001). The regression analysis (i.e., prediction of one set of item difficulties from the other set of difficulty scores) yielded F = 332.07 ($p < .0001$). These figures indicate that the two sets of item difficulties are very highly correlated. One would expect this to be so, given that the tests used identical questions (although presented in a slightly different order) and were given to two homogeneous samples of testees.

Test information functions. The test information functions are presented in Figures 1 and 2. The vertical axis represents the amount of "information" provided by each test. The horizontal axis represents ability levels (i.e., the presence or absence of the latent trait) of testees. Recall that the information function is a measure of the precision of ability estimation at each ability level. That is, the standard error of measurement has been calculated for each ability level, not the test in its entirety, as is the case with classical test models. As can be seen, the two information

functions are nearly identical. The curves are somewhat bell-shaped, not flat (as had been anticipated).

Conclusions:

Use of the Rasch model. Given the results of the initial factor analyses, it would appear that the data fit the assumption of test unidimensionality, as required by the Rasch model analysis. Regarding the other two assumptions--that of non-existent guessing and equal item discrimination levels--it seems reasonable to assume that guessing and/or cheating was minimized during the test administration by staggering the test forms throughout the testee group and by careful preparation of test question "foils". Also, the standard item analysis of the items on the two test forms revealed that the PBIS (point biserial) measures (albeit a somewhat limited statistic given its sample dependence) for the majority of items ranged from +.25 to +.35. One could argue that, for the purposes of this investigation, the majority of items can be considered as having "equal" discrimination levels. In summary, it would seem that the data fits the assumptions for data to be used in a Rasch logistic model analysis.

Research questions. At the outset of the analysis, two research questions were raised:

1. Could the Rasch model recover "known" item parameters?

2. Were the shapes of the tests' information curves flat, representing equal precision of measurement across all ability levels?

As regards the first research question, it is clear from the results of the analysis that the Rasch model was able to recover known item parameters ("known" in the sense that they should be equal). The Rasch item difficulties for the two test forms were

very highly correlated (r = .907, p = .001). In addition, the test information functions were nearly identical.

Regarding the second research question, it is apparent that the shapes of the two test information functions are not flat. Rather, they are bell-shaped, indicating that more accurate information is being provided over the mid-range of the ability continuum than at the extreme ends of the scale. This is not advantageous in a test of this kind in that, ideally, one desires equal information accuracy accross all ability levels in an achievement test over general course content. Loss of some information at the extreme ends of the ability continuum are not, however, devastating to the tests' ability to fairly assess the majority of testees (i.e., those that fall into the mid-range of ability).

Implications for improvements on the tests. Generally speaking, one can use the information provided by the Rasch logistic model analysis in two ways: a) evaluate the easiness/difficulty of the test items, and b) evaluate the shape of the test information curve.

Given the results of the analysis presented in this paper, it seems that that tests have a fairly equal balance of "easy" and "difficult" items. This is typical of tests of this kind and is to the tests' advantage. If the test designer required a more difficult test, s/he could discard some of the "easy" items in favor of more items in the "difficult" rance, or vice versa. This is a question of designer discretion, but either option is available with the information provided by the Rasch analysis.

Regarding the evaluation of the shapes of the tests' information functions, it is clear that the functions are bell-shaped, not flat (as had been anticipated). However, the tests' greatest accuracy is in the mid-range of testee ability, with less accuracy at the extremes of the ability continuum. For this reason, the test can be said to be providing accurate information for the majority of testees (i.e., those with moderate ability). In light of this, the test designer might be satisfied with the information provided by the tests as they are.

However, if the test designer committed to improving the estimation of ability at the extremes of the ability continuum, item information curves provided by the latent trait model analysis would allow him/her to do so. Lord (1977) suggests this four step process:

1. Decide on the shape desired for the test information function (e.g., flat, as opposed to bell-shaped). This is the "target information curve".

2. Select items with item information curves that will fill the hard-to-fill areas under the target information curve.

3. Cumulatively add up the item information curves, obtaining at all times the information curve for the partial test composed of items already selected.

4. Continue (revising the revisions, if necessary) until the area under the target information curve is filled up satisfactorily.

So, if the test designer wanted a flatter curve, s/he could use the process described above to do so. The finished product would be a test information function that would provide accurate estimates of ability at all ability levels, regardless of the characteristics of the sample of testees used in the analysis.

Discussion

In this paper, a brief and non-technical introduction to latent trait models was presented. In addition, an argument was made that due to the properties of objectivity and information, as well as their ability to identify biased test items, latent trait models are superior to classical test models. Also, due to its ease of application and the usefulness of the test statistics generated, the one parameter, or Rasch, logistic model was presented as the most desirable of the logistic models. Finally, an example was given of the use of the Rasch model in the context of communication education.

It should be clear that latent trait models can provide helpful answers to the questions raised at the outset of this paper. For example, use of latent trait models provides information regarding all of the following: difficulty of test items, unidimensionality of the test (i.e., in terms of ability tested), identification of test item bias, relative contribution of information from any given test item, and accuracy of information provided at each level of testee ability. Latent trait models are a valuable "means" to the ideal "end" of testing: that is, devising test instruments that will provide stable and accurate assessments of examinee performance. It's time for communication professionals to take advantage of this new educational tool.

References

Allen, M. J. and Yen, W. M. Introduction to measurement theory. Monterey, CA: Brooks/Cole Publishing Company, 1979.

Hambleton, R. K. and Cook, L. L. Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 1977, 14, 75-96.

Lord, F. M. Practical applications of item characteristic curve theory. Journal of Educational Measurement, 1977, 14, 117-138.

Lord, F. M. Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, 1980.

Rasch, G. An item analysis which takes individual differences into account. British Journal of Mathematical and Statistical Psychology, 1966, 19, 49-57.

Wright, B. D. Sample-free test calibration and person measurement. Paper presented to the 1967 Invitational Conference on Testing Problems.

Wright, B. D. and Stone, M. H. Best test design. Chicago: Mesa Press, 1979.

Table 1

Rasch Item Difficulties

Test Forms A and B

Item #	Form A	Form B	Item #	Form A	Form B
1	+2.686	+ .365	41	-1.592	-2.292
2	- .191	- .547	42	-1.153	- .987
3	- .692	-1.257	43	+ .914	+ .526
4	+ .963	+ .488	44	+ .218	- .112
5	+1.121	+1.669	45	+2.067	+1.855
6	+2.067	+2.705	46	-1.222	-1.302
7	-1.130	-1.122	47	+2.496	+3.238
8	+1.434	+1.669	48	- .109	- .523
9	+ .690	+ .685	49	- .271	- .547
10	- .349	- .450	50	+1.366	+1.855
11	+1.850	+1.960	51*	-3.079	+ .060
12	- .573	- .425	52	- .191	- .324
13	- .081	+ .246	53	+1.121	+1.855
14	-1.836	-1.940	54	+ .690	+ .685
15	1.267	-1.415	55	- .164	+ .030
16	-1.085	-1.167	56	- .218	- .324
17	+ .250	+ .120	57	-2.180	-2.412
18	- .245	- .299	58	-1.616	-1.553
19	- .245	- .139	59	+ .732	+1.010
20	+ .124	+ .182	60	+ .094	+ .182
21	+2.192	+2.203	61	+1.667	+1.299
22	- .645	- .350	62	- .597	-1.032
23	+ .776	+ .862	63	+ .820	-3.112
24	- .499	- .028	64	+ .124	- .028
25	+1.507	+ .959	65	+ .155	+ .030
26	+ .283	+ .246	66	- .081	- .084
27	- .191	+ .030	67	- .524	- .690
28	- .716	- .965	68	- .499	- .220
29	+ .064	+ .381	69	+ .094	+ .910
30	- .856	- .919	70	-2.611	-2.745
31	- .191	- .499	71	- .218	- .220
32	+ .690	+1.119	72	-3.079	-2.820
33	+1.366	+1.434	73	+ .457	+ .279
34	+ .649	+1.508	74	-1.522	-1.647
35	+ .866	+ .910	75	-1.569	-1.553
36	- .716	-1.032			
37	- .971	-1.055			
38	+2.192	+2.511			
39	+ .690	+ .346			
40	- .499	- .112			

*Item was keyed improperly. Results for this item are, therefore, invalid.

Figure 1. Test information function: Test form A

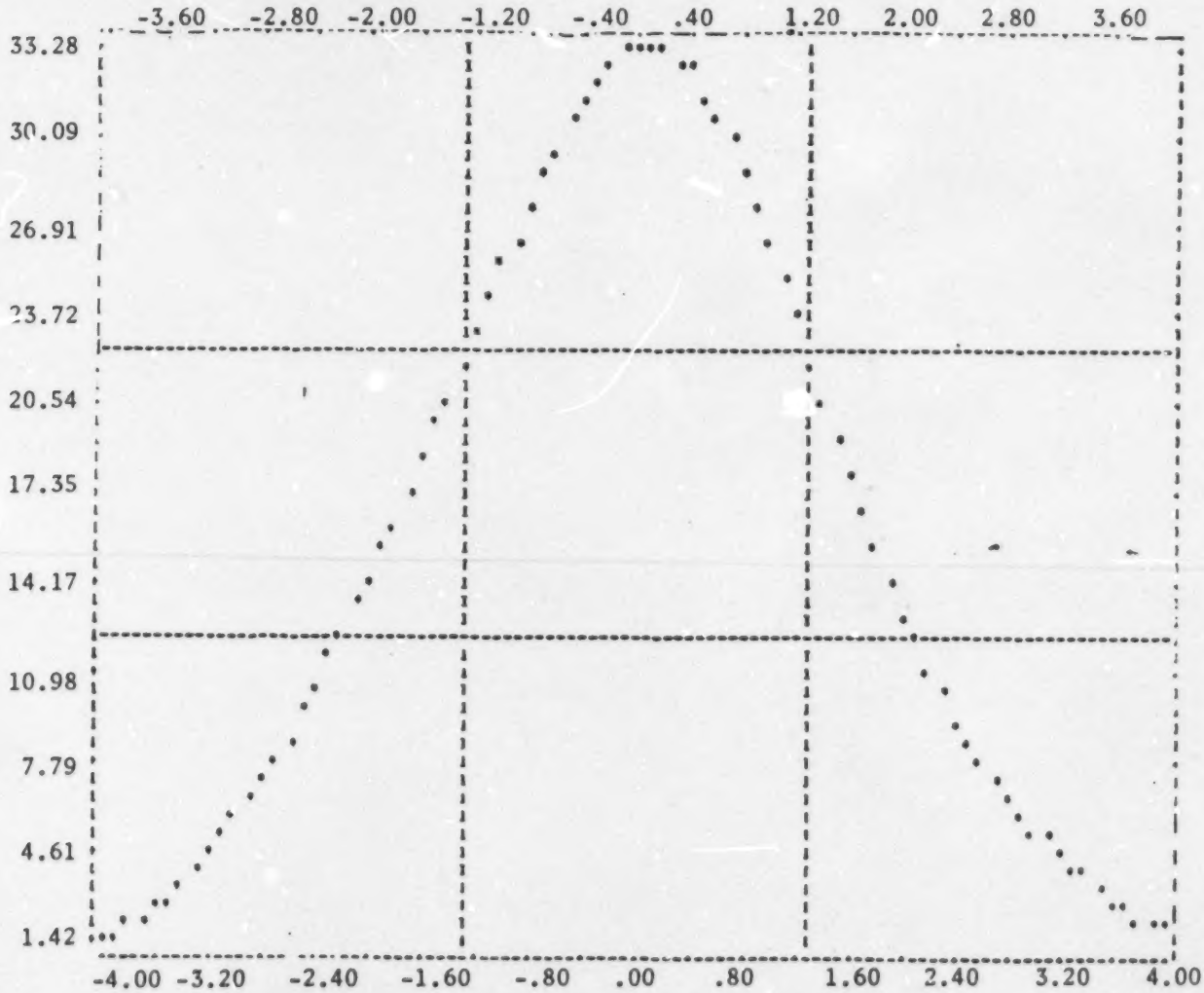

Figure 2. Test information function: Test form B